

Übungen zum Bioinformatik-Tutorium

Blatt 3

Termin: Donnerstag, 29.06.16, 12 Uhr

Hinweis

Die Aufgaben sollen mit *R* gelöst werden. Dafür können Sie die veraltete, vorinstallierte Version von Rstudio, den interaktiven Modus in der Shell, oder einen beliebigen Texteditor verwenden. Versuchen Sie Aufgabe 1-a und 1-b mit der Shell zu lösen.

1. Datei Handling

- (a) Legen Sie sich einen Ordner an, um die folgende Aufgabe darin zu lösen.
- (b) Kopieren Sie die Datei `/home/proj/tutorium_bioinformatik/yeastPheromoneResponseExpr.tsv` in ihren angelegten Ordner.
- (c) Schreiben Sie nun **ein** Skript, dass ...

... die Datei einliest und den Inhalt in einer Matrix abspeichert. Dabei soll die erste Zeile der Datei als Kopfzeile dienen, die die Spalten beschreibt. Überprüfen Sie, ob die Matrix korrekt abgespeichert wurde.

... die Zeilen- und Spaltenanzahl auf der Konsole ausgibt.

... die eindeutigen Namen, der Proteine aus der ersten Spalte, sortiert ausgibt. Wie viele sind das?

... die Werte der Spalten 2-7 auf der Konsole ausgibt, die in der ersten Spalte der Matrix den Namen YAL020C, YAL048C, YBR041W oder YDR109C gespeichert haben.

... den Inhalt von der eingelesenen Matrix in die Datei `yeastPherRespExpr.tsv` abspeichert. Überprüfen Sie, ob die erzeugte Datei der Ursprungsdatei entspricht.

2. Plotten

Die Datei aus Aufgabe 1 enthält Expressionsdaten eines two-channel Microarray Experiments über eine Zeitspanne (0min, 15min, 30min ...). In einem two-channel Microarray werden zwei RNA Proben unterschiedlich gelabelt und gleichzeitig gemessen, so dass man einen fold change erhält (relative Messung: z.B. doppelt so viel Probe 1 als Probe 2). Die Werte aus diesen Messungen entsprechen den Werten in der Datei. In diesem Experiment wurden Hefezellen mit einem Pheromon (Probe 1) und einer Kontrolle (Probe 2) behandelt. Dabei wurden zu verschiedenen Zeitpunkten Messungen durchgeführt. Die erste Messung wurde direkt nach der Stimuluszugabe gemacht, so dass noch kein Effekt erwartet wird. Diese Messung kann benutzt werden um systematische Effekte zwischen den beiden Samples zu reduzieren. Ihre Aufgabe ist es nun, die fold changes der einzelnen Zeitpunkte (> 0) im Bezug auf den Zeitpunkt 0 zu bestimmen. Hier lässt sich ein fold change wie folgt berechnen

$$fcs = t_i - t_0$$

wobei $i > 0$ ist. Speichern Sie dabei die berechneten Werte in einer neuen Matrix ab. Achten Sie darauf, dass die Ergebnisse spaltenweise abgespeichert werden. Diese Matrix soll dann verwendet werden, um die folgenden Plots zu erzeugen und abzuspeichern:

- Boxplot. Sieht man Veränderungen über die unterschiedlichen Zeiträume?
- Heatmap.

Hinweis:

Schleifen und Vektoroperationen eignen sich zum Lösen der Aufgabe.

Erweiterung:

Erzeugen Sie einen Plot (x,y), indem man die fold changes der ersten 10 Gene über die unterschiedlichen Zeiträume sieht. Dabei sollen alle Linien in dem selben Plot sein.